

WHAT IS CLAIMED IS:

1. A speech detection system, comprising:  
an audio microphone outputting a microphone signal based on a sensed audio input;  
a speech sensor outputting a sensor signal based on a non-audio input generated by speech action; and  
a speech detector component outputting a speech detection signal indicative of whether a user is speaking based on the sensor signal.
2. The speech detection system of claim 1 wherein the speech detector component outputs the speech detection signal based on a first characteristic of the sensor signal and based on the microphone signal.
3. The speech detection system of claim 2 wherein the first characteristic of the sensor signal has a first level when the user is speaking and a second level when the user is not speaking and wherein the speech detector component outputs the speech detection signal based on a level of the first characteristic of the sensor signal relative to a baseline level of the first characteristic that comprises a predetermined one of the first and second levels of the characteristic.

4. The speech detection system of claim 3 wherein the baseline level is calculated based on a level of the first characteristic over a time period.

5. The speech detection system of claim 4 wherein the baseline level is calculated by averaging the level of the first characteristic over the time period.

6. The speech detection system of claim 4 wherein the baseline level is recalculated intermittently during operation of the speech detection system.

7. The speech detection system of claim 6 wherein the baseline level is recalculated periodically to represent the level of the first characteristic over a revolving time window.

8. The speech detection system of claim 6 wherein the speech detection component outputs the speech detection signal based on a comparison of the level of the first characteristic of the sensor signal to the baseline level, and wherein the comparison is performed periodically.

9. The speech detection system of claim 9 wherein the comparison is performed more frequently than the baseline level is recalculated.

10. The speech detection system of claim 1 wherein the audio microphone and the speech sensor are mounted to a headset.

11. A speech recognition system, comprising:  
a speech detection system comprising:  
an audio microphone outputting a microphone signal based on a sensed audio input;  
a speech sensor outputting a sensor signal based on a non-audio input generated by speech action; and  
a speech detector component outputting a speech detection signal indicative of whether a user is speaking based on the microphone signal and the sensor signal; and  
a speech recognition engine providing a recognition output indicative of speech in the sensed audio input based on the microphone signal and the speech detection signal.

12. The speech recognition system of claim 11 wherein the speech detector component calculates the speech detection signal as a speech detection measure, indicative of a probability that the user is speaking.

13. The speech recognition system of claim 12 wherein the speech detector component combines the

speech detection measure with the microphone signal to generate a combined signal.

14. The speech recognition system of claim 13 wherein the speech recognition engine generates the recognition output based on the combined signal.

15. The speech recognition system of claim 14 wherein the speech detection measure comprises a probability that the user is speaking.

16. The speech recognition system of claim 15 wherein the combined signal comprises a product of the probability and the microphone signal.

17. The speech recognition system of claim 11 wherein the audio microphone and the speech sensor being mounted on a headset.

18. A method of detecting speech, comprising:  
generating a first signal, indicative of an audio input, with an audio microphone;  
generating a second signal indicative of facial movement of a user, sensed by a facial movement sensor; and  
detecting whether the user is speaking based on the first and second signals.

19. The method of claim 18 wherein generating the second signal comprises:

sensing vibration of one of the user's jaw and neck.

20. The method of claim 18 wherein generating the second signal comprises:

sensing an image indicative of movement of the user's mouth.

21. The method of claim 18 and further comprising:

providing a speech detection signal based on detecting whether the user is speaking.

22. The method of claim 21 and further comprising:

recognizing speech based on the first signal and the speech detection signal.

23. The method of claim 22 wherein recognizing speech comprises:

increasing a likelihood that speech is recognized if the speech detection signal indicates that the user is speaking; and  
decreasing a likelihood that speech is recognized if the speech detection signal indicates that the speaker is not speaking.